Offline Reinforcement Learning with Fisher Divergence Critic Regularization

Year: 2021 | Citations: 335 | Authors: Ilya Kostrikov, Jonathan Tompson, R. Fergus, Ofir Nachum

---

Abstract

Many modern approaches to offline Reinforcement Learning (RL) utilize behavior regularization, typically augmenting a model-free actor critic algorithm with a penalty measuring divergence of the policy from the offline data. In this work, we propose an alternative approach to encouraging the learned policy to stay close to the data, namely parameterizing the critic as the log-behavior-policy, which generated the offline data, plus a state-action value offset term, which can be learned using a neural network. Behavior regularization then corresponds to an appropriate regularizer on the offset term. We propose using a gradient penalty regularizer for the offset term and demonstrate its equivalence to Fisher divergence regularization, suggesting connections to the score matching and generative energy-based model literature. We thus term our resulting algorithm Fisher-BRC (Behavior Regularized Critic). On standard offline RL benchmarks, Fisher-BRC achieves both improved performance and faster convergence over existing state-of-the-art methods.